

THE DIRICHLET PROCESS WITH LARGE CONCENTRATION PARAMETER

LUAI AL LABADI AND MAHMOUD ZAREPOUR

Abstract: Ferguson's Dirichlet process plays an important role in nonparametric Bayesian inference. Let P_a be the Dirichlet process in \mathbb{R} with a base probability measure H and a concentration parameter $a > 0$. In this paper, we show that $\sqrt{a}(P_a((-\infty, t]) - H((-\infty, t]))$ converges to a certain Brownian bridge as $a \rightarrow \infty$. We also derive a certain Glivenko-Cantelli theorem for the Dirichlet process. Using the functional delta method, the weak convergence of the quantile process is also obtained. A large concentration parameter occurs when a statistician puts too much emphasize on his/her prior guess. This scenario also happens when the sample size is large and the posterior is used to make inference.

Key words and phrases: Bayesian nonparametric, Brownian bridge, Dirichlet process, quantile process, weak convergence.

1 Introduction

In nonparametric Bayesian inference, we need to place a prior on an infinite dimensional space such as the space of probability measures. Ferguson (1973) used a Dirichlet process (a normalized gamma process) as a prior on this space. For $k \geq 2$, we say that the random vector (Y_1, \dots, Y_k) has the Dirichlet distribution with parameters (a_1, \dots, a_k) , where $a_i > 0$ for all i , if it has the joint probability density function

$$f(y_1, \dots, y_k) = \frac{\Gamma\left(\sum_{i=1}^k a_i\right)}{\prod_{i=1}^k \Gamma(a_i)} \prod_{i=1}^k y_i^{a_i-1} I_{\mathbb{S}}(y_1, \dots, y_k),$$

where $\mathbb{S} = \{(y_1, \dots, y_k) : y_i \geq 0, \sum_{i=1}^k y_i = 1\}$ and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, $x > 0$. We denote by $D(a_1, \dots, a_k)$ the Dirichlet distribution with parameters a_1, \dots, a_k .

The Dirichlet process was defined in Ferguson (1973) as follows: let $(\mathfrak{X}, \mathcal{A})$ be an arbitrary measurable space and H be a probability measure on $(\mathfrak{X}, \mathcal{A})$. Let $a > 0$ be arbitrary. A random probability measure $P_a = \{P_a(A)\}_{A \in \mathcal{A}}$ is called a Dirichlet process on $(\mathfrak{X}, \mathcal{A})$ with parameters a

and H , if for any finite measurable partition $\{A_1, \dots, A_k\}$ of \mathfrak{X} , the joint distribution of the vector $(P_a(A_1), \dots, P_a(A_k))$ has the Dirichlet distribution with parameters $(aH(A_1), \dots, aH(A_k))$. The subscript a is added since in the forthcoming sections we will study the asymptotic behavior of the random probability measure P_a for large values of a . We assume that if $H(A_j) = 0$, then $P_a(A_j) = 0$ with probability one. We write $P \sim \text{DP}(a, H)$ to denote the Dirichlet process with parameters a and H . Throughout this paper, we use the same letter for the probability measure and its corresponding cumulative distribution function, i.e. $P_a(t) = P_a((-\infty, t])$ and $H(t) = H((-\infty, t])$. We also assume that the cumulative distribution function H is continuous.

For any $A \in \mathcal{A}$, $P_a(A)$ has a Beta distribution with parameters $aH(A)$ and $a(1 - H(A))$. Thus,

$$E(P_a(A)) = H(A) \quad \text{and} \quad \text{Var}(P_a(A)) = \frac{H(A)(1 - H(A))}{1 + a}. \quad (1.1)$$

Furthermore, for any two sets A_i and $A_j \in \mathcal{A}$, it follows from the properties of a Dirichlet distribution that (Wilks 1963, page 177)

$$E(P_a(A_i)P_a(A_j)) = \frac{a}{1 + a}H(A_i)H(A_j) \quad (1.2)$$

The probability measure H is called the base measure of P_a . Clearly, from (1.1), H plays the role of the center of the process, while a can be viewed as the concentration parameter. The larger a is, the more likely it is that the realization of P is close to H . Specifically, for any fixed set $A \in \mathcal{A}$ and $\epsilon > 0$, we have $P_a(A) \xrightarrow{P} H(A)$ as $a \rightarrow \infty$ since

$$\Pr\{|P_a(A) - H(A)| > \epsilon\} \leq \frac{H(A)(1 - H(A))}{\epsilon^2(1 + a)}.$$

In this paper, “ \xrightarrow{P} ” denotes the convergence in probability.

An attractive property of the Dirichlet process is its conjugacy property. That is, if X_1, \dots, X_n is a random sample from $P_a \sim \text{DP}(a, H)$, then the posterior distribution of P_a given X_1, \dots, X_n coincides with the distribution of the Dirichlet process with parameter measure a^*H^* , where

$$a^* = a + n \quad \text{and} \quad H^* = \frac{a}{a + n}H + \frac{n}{a + n} \frac{\sum_{k=1}^n \delta_{X_k}}{n}. \quad (1.3)$$

Here and throughout the paper δ_X denotes the Dirac measure at X , i.e. $\delta_X(A) = 1$ if $X \in A$ and 0 otherwise.

Notice that the posterior base distribution H^* is a convex combination of the base distribution and the empirical distribution. The weight associated with the prior base distribution H is proportional to a , giving another reason to call a the concentration parameter. The weight associated with the empirical distribution is proportional to the number of observations n . The posterior base distribution H^* approaches the prior base measure H for large values of a . On the other hand, for small values of a , H^* is close to the empirical process.

The Dirichlet process has the following series representation:

$$P_a(\cdot) = \sum_{i=1}^{\infty} J_i \delta_{\theta_i}(\cdot), \quad (1.4)$$

where $(\theta_i)_{i \geq 1}$ is a sequence of independent and identically distributed (i.i.d.) random variables with common distribution H and $(J_i)_{i \geq 1}$ are random variables chosen to be independent of $(\theta_i)_{i \geq 1}$ and such $0 \leq J_i \leq 1$ and $\sum_{i=1}^{\infty} J_i = 1$ almost surely. For several representations for $(J_i)_{i \geq 1}$, see, for example, Ferguson, Phadia, and Tiwari (1992). It follows from (1.4) that any realization of the Dirichlet process must be a discrete probability measure.

Sethuraman and Tiwari (1982) studied the convergence and tightness of Dirichlet processes as the parameters are allowed to converge in a certain sense. They showed that as the concentration parameter $a \rightarrow 0$, the Dirichlet process converges to a degenerate probability measure at a particular point in \mathfrak{X} randomly chosen from H .

Let \mathcal{S} be a collection of Borel sets in \mathbb{R} . For large values of the concentration parameter a , we study the weak convergence of the centralized and scaled Dirichlet process defined by

$$D_a(S) = \sqrt{a} (P_a(S) - H(S)), \quad S \in \mathcal{S}. \quad (1.5)$$

We also derive the limiting distribution of the Dirichlet quantile process

$$Q_a = \sqrt{a} (P_a^{-1} - H^{-1}), \quad (1.6)$$

where in general the inverse of a distribution function F is given by $F^{-1}(t) = \inf \{x : F(x) \geq t\}$. Moreover, a certain Glivenko-Cantelli theorem for the Dirichlet process for large values of concen-

tration parameter is obtained.

For the Dirichlet posterior processes with parameters given in (1.3), the concentration parameter $a^* \rightarrow \infty$ whenever $n \rightarrow \infty$ (n is the sample size). Lo (1987) studied completely the behavior of the process

$$d_{n,a}(t) = \sqrt{n} (P_{n,a}^*(t) - F_n(t)), \quad t \in \mathbb{R},$$

as the sample size n gets large, where $P_{n,a}^*$ is the posterior of the Dirichlet process P_a given the data and F_n is the empirical distribution function. Using this result, Lo (1987) gave an asymptotic justification of the use of Bayesian bootstrap and provided large sample Bayesian bootstrap probability intervals for the mean, the variance, and bands for the distributions.

2 Asymptotic Properties of the Dirichlet process

In this section, we study the asymptotic properties of P_a as $a \rightarrow \infty$, where $P_a \sim DP(a, H)$. Since H is strictly increasing, we have

$$\theta_i < t \text{ if and only if } H(\theta_i) < H(t).$$

Thus,

$$P_a((-\infty, t]) = \sum_{i=1}^{\infty} J_i \delta_{\theta_i}((-\infty, t]) \stackrel{d}{=} \sum_{i=1}^{\infty} J_i \delta_{H(\theta_i)}((0, H(t)]).$$

Throughout this paper, “ $\stackrel{d}{=}$ ” denotes equality in distribution. Since $(\theta_i)_{i \geq 1}$ is a sequence of i.i.d. random variables with continuous distribution H , for $i \geq 1$, $U_i \stackrel{d}{=} H(\theta_i)$ where $\{U_i\}_{i \geq 1}$ is a sequence of i.i.d. random variables with a uniform distribution on $[0, 1]$. Hence,

$$P_a((-\infty, t]) \stackrel{d}{=} \sum_{i=1}^{\infty} J_i \delta_{U_i}((0, H(t)]).$$

Therefore, without loss of generality, we only consider the case when $H(t) = t$ (i.e., $(\theta_i)_{i \geq 1}$ is a sequence of i.i.d. random variables with uniform distribution on $[0, 1]$). Hence, the process in (1.5) reduces to

$$D_a(S) = \sqrt{a} (P_a(S) - \lambda(S)), \quad (2.1)$$

where λ is the Lebesgue measure on $[0, 1]$. Hereafter, unless otherwise stated, $P_a \sim DP(a, \lambda)$, where λ is the Lebesgue measure on $\mathfrak{X} = [0, 1]$.

We now recall the definition of a Brownian bridge indexed by \mathcal{S} . A Gaussian process $\{B_\lambda(S) : S \in \mathcal{S}\}$ is called a Brownian bridge if

$$E[B_\lambda(S)] = 0 \quad \text{and} \quad \text{Cov}(B_\lambda(S_i), B_\lambda(S_j)) = \lambda(S_i \cap S_j) - \lambda(S_i)\lambda(S_j), \quad (2.2)$$

where $S, S_i, S_j \in \mathcal{S}$ (Massart 1989).

The next lemma gives the limiting distribution of the process (2.1) for any finite Borel sets $S_1, \dots, S_k \in \mathcal{S}$. The proof of the lemma for $k = 2$ is given in the appendix and it can be generalized easily to the case of arbitrary k . In this paper, “ \xrightarrow{d} ” denotes the convergence in distribution.

Lemma 1. *Let D_a be as defined in (2.1). Then, as $a \rightarrow \infty$, for any fixed sets $S_1, \dots, S_k \in \mathcal{S}$,*

$$(D_a(S_1), D_a(S_2), \dots, D_a(S_k)) \xrightarrow{d} (B_\lambda(S_1), B_\lambda(S_2), \dots, B_\lambda(S_k)),$$

where B_λ is the Brownian bridge indexed by \mathcal{S} with the mean and the covariance structure as given in (2.2).

Remark 1. The convergence obtained in Lemma 1 is called convergence in total variation. This type of convergence is stronger than convergence in distribution (Billingsley 1999, page 29).

Remark 2. It follows from Lemma 1 that, for any fixed Borel set $S \in \mathcal{S}$,

$$D_a(S) = \sqrt{a}(P_a(S) - \lambda(S)) \xrightarrow{d} B_\lambda(S),$$

where $B_\lambda(S)$ is distributed as $N(0, \lambda(S)(1 - \lambda(S)))$.

Lemma 1 proves that the finite-dimensional distributions of the process D_a converge to the corresponding finite-dimensional distribution of B_λ . The next theorem shows that the process D_a converges to the process B_λ on $D[0, 1]$ with respect to the Skorokhod topology.

Theorem 1. *Let D_a be as defined in (2.1). Then, as $a \rightarrow \infty$, we have:*

$$\sqrt{a}(P_a(\cdot) - \lambda(\cdot)) \xrightarrow{d} B_\lambda(\cdot)$$

on $D[0, 1]$ with respect to the Skorokhod topology, where B_λ is a Brownian bridge.

Proof. From Lemma 1 and Theorem 13.5 of Billingsley (1999) we need only to prove that for any $0 \leq t_1 \leq t \leq t_2 \leq 1$,

$$E \left[|P_a(t) - P_a(t_1)|^{2\beta} |P_a(t_2) - P_a(t)|^{2\beta} \right] \leq |F(t_2) - F(t_1)|^{2\alpha},$$

for some $\beta \geq 0$, $\alpha > 1/2$, and a nondecreasing continuous function F on $[0, 1]$. Take $\beta = 1/2$, $\alpha = 1$, and $F(t) = t$ to show that:

$$E [(P_a(t) - P_a(t_1)) (P_a(t_2) - P_a(t))] \leq \frac{a}{a+1} (t_2 - t_1)^2. \quad (2.3)$$

Observe that,

$$(P_a(t) - P_a(t_1), P_a(t_2) - P_a(t)) \sim D(a\lambda(t_1, t], a\lambda(t, t_2], a(1 - \lambda(t_1, t] - \lambda(t, t_2]))$$

From (1.2) we have:

$$\begin{aligned} E [(P_a(t) - P_a(t_1)) (P_a(t_2) - P_a(t))] &= \frac{a}{a+1} \lambda(t_1, t] \lambda(t, t_2] \\ &= \frac{a}{a+1} (t - t_1) (t_2 - t). \end{aligned}$$

Since $t_1 \leq t \leq t_2$, (2.3) follows. This completes the proof of the theorem. \square

As in Ferguson (1973), under the squared error loss and Dirichlet prior, the no data estimate (or the posterior estimate) for the distribution is the prior distribution H . Under the absolute deviation loss, the estimate will be the median of the Dirichlet process with the prior distribution of H . Therefore, the Dirichlet quantile process plays a role in estimation. The following corollary derives the asymptotic behavior of the Dirichlet quantile process defined by (1.6) when the concentration parameter a is large.

Corollary 1. *Let $0 < p < q < 1$, and H be a continuous function with positive derivative h on the interval $[H^{-1}(p) - \epsilon, H^{-1}(q) + \epsilon]$ for some $\epsilon > 0$. Let Q_a be the Dirichlet quantile process defined in (1.6), where $P_a \sim DP(a, H)$. Then, as $a \rightarrow \infty$, we have:*

$$Q_a(\cdot) \xrightarrow{d} -\frac{B_\lambda(\cdot)}{h(H^{-1}(\cdot))} = Q(\cdot),$$

in $D[p, q]$. That is, the limiting process is a Gaussian process with zero-mean and covariance func-

tion

$$\text{Cov}(Q(S_i), Q(S_j)) = \frac{\lambda(S_i \cap S_j) - \lambda(S_i)\lambda(S_j)}{h(H^{-1}(S_i))h(H^{-1}(S_j))}, \quad S_i, S_j \in \mathcal{S}.$$

Proof. By Theorem 1 the process $\sqrt{a}(P_a - H)$ converges in distribution to the process $B_H = B_\lambda(H) = B_\lambda \circ H$. Almost all sample paths of the limiting process are continuous on the interval $[H^{-1}(p) - \epsilon, H^{-1}(q) + \epsilon]$. By Lemma 3.9.23. page 386 of Van der Vaart and Wellner (1996), the inverse map $H \mapsto H^{-1}$ is Hadamard-differentiable at H tangentially to the subspace of functions that are continuous on this interval. By the functional delta method (Theorem 3.9.4 page 374 of Van der Vaart and Wellner (1996)) we have

$$Q_a(\cdot) \xrightarrow{d} -\frac{B_\lambda \circ H \circ H^{-1}(\cdot)}{h(H^{-1}(\cdot))} = -\frac{B_\lambda(\cdot)}{h(H^{-1}(\cdot))}$$

in $D[p, q]$. This completes the proof of the corollary. \square

Remark 3. Paralleling Remark 1 of Bickel and Freedman (1981), if $H^{-1}(0+) > -\infty$ and $H^{-1}(1) < \infty$ and h is continuous on $[H^{-1}(0+), H^{-1}(1)]$, the conclusion of the corollary holds in $D[H^{-1}(0+), H^{-1}(1)]$. For example, if H is a uniform distribution on $[0, 1]$, then the convergence holds in $D[0, 1]$. More generally, we may have one end of the support finite and the other infinite and a modified form of Corollary 1 still holds. Also from the result of Theorem 1, we can derive asymptotic properties of any Hadamard-differentiable functional of the $DP(a, H)$ as $a \rightarrow \infty$.

Example 1 (Median). Let M_a be the median of P_a and m be the median of H (i.e. $P_a^{-1}(0.5) = M_a$ and $H^{-1}(0.5) = m$). From Corollary 1 we have:

$$\sqrt{a}(M_a - m) \xrightarrow{d} N\left(0, \frac{1}{4h^2(m)}\right),$$

where $h = H'$. Note that, the asymptotic distribution of the median for Dirichlet process coincide with that of the sample median.

Example 2 (Interquantile Range). Similar to Example 1, let $IQR = Q_{3,a} - Q_{1,a}$, where $Q_{3,a}$ and $Q_{1,a}$ are the third and the first quartiles of P_a (i.e. $P_a^{-1}(0.75) = Q_{3,a}$ and $P_a^{-1}(0.25) = Q_{1,a}$). Let q_3 and q_1 be the third and the first quartiles of H . From Corollary 1, a simple calculation shows

$$\sqrt{a}(IQR - (q_3 - q_1)) \xrightarrow{d} N\left(0, \frac{3}{h^2(q_3)} + \frac{3}{16h^2(q_1)} - \frac{2}{h(q_1)h(q_3)}\right).$$

This gives with the asymptotic distribution of the sample interquartile range.

In the next theorem we establish the Glivenko-Cantelli theorem for the Dirichlet process. In this paper, “ $\xrightarrow{a.s.}$ ” denotes the almost sure convergence.

Theorem 2. *Let $P_a \sim DP(a, H)$. Then,*

$$\sup_{x \in \mathbb{R}} |P_a(x) - H(x)| \xrightarrow{a.s.} 0,$$

as $a \rightarrow \infty$.

Proof. From Donoho and Liu (1988),

$$\frac{(\sup_x |P_a(x) - H(x)|)^{3/2}}{3^{1/2}} \leq \int_{-\infty}^{\infty} (P_a(x) - H(x))^2 dH(x). \quad (2.4)$$

Notice that $P_a(x) \xrightarrow{p} H(x)$, as $a \rightarrow \infty$, and $(P_a(x) - H(x))^2$ is dominated by 1. Thus, by the dominated convergence theorem (which remains valid for convergence in probability (Royden 1968, page 92)), we obtain that the right hand side of (2.4) converges to zero. \square

When the concentration parameter is large, the Dirichlet process and its corresponding quantile process share many asymptotic properties with the empirical process and the quantile process.

3 Acknowledgments.

The authors would like to thank Professor Raluca Balan for her helpful comments and suggestions. The research of the authors are supported by research grants from the **Natural Sciences and Engineering Research Council of Canada (NSERC)**.

REFERENCES

- [1] Bickel, P. J., and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196-1217.
- [2] Billingsley, P. (1999). *Convergence of Probability Measures*, third ed. John Wiley & Sons, Inc.

- [3] Donoho, D. L., and Liu, R. C. (1988). Pathologies of some minimum distance estimator. *Ann. Statist.*, **2**, 587-608.
- [4] Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *Ann. Statist.*, **1**, 209-230.
- [5] Ferguson, T. S., Phadia, E. G. and Tiwari, R. C. (1992). Bayesian nonparametric inference. In *current Issues in Statistical Inference: Essays in Honor of D. Basu*, Ed. M. Ghosh and P. K. Pathak, institute of Mathematical Statistics, pp. 127-150.
- [6] Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.*, **15**, 360-375.
- [7] Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via K.M.T. constructions. *Ann. Probab.*, **17**, 266-291.
- [8] Royden, H. L. (1968). *Real Analysis*, 3rd ed., New York: Macmillan.
- [9] Sethuraman, J. and Tiwari, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter, *Statistical Decision Theory and Related Topics III*, **2**, 305-315.
- [10] Van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer.
- [11] Wilkes, S. S. (1963). *Mathematical Statistics*. John Wiley & Sons, Inc.

Luai Al-Labadi is a PhD Student, Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada,

E-mail: (lalla046@uottawa.ca)

Mahmoud Zarepour is Associate Professor, Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada,

E-mail: (zarepour@uottawa.ca)

Appendix: Proof of Lemma 1 for $k = 2$

Assume that $S_1 \cap S_2 = \emptyset$. (The general case when S_1 and S_2 are not necessarily disjoint follows from the continuous mapping theorem). Note that

$$(P_a(S_1), P_a(S_2), 1 - P_a(S_1) - P_a(S_2)) \sim D(a\lambda(S_1), a\lambda(S_2), a(1 - \lambda(S_1) - \lambda(S_2)))$$

Set $X_{i,a} = P_a(S_i)$ and $l_i = \lambda(S_i)$, $i = 1, 2$. Thus, the joint density function of $P_{1,a}$ and $P_{2,a}$ is:

$$f_{X_{1,a}, X_{2,a}}(x_1, x_2) = \frac{\Gamma(a)}{\Gamma(al_1)\Gamma(al_2)\Gamma(a(1 - l_1 - l_2))} x_1^{al_1-1} x_2^{al_2-1} (1 - x_1 - x_2)^{a(1-l_1-l_2)-1}.$$

The joint probability density function of $D_{1,a} = \sqrt{a}(X_{1,a} - l_1) = \sqrt{a}(P_a(S_1) - \lambda(S_1))$ and $D_{2,a} = \sqrt{a}(X_{2,a} - l_2) = \sqrt{a}(P_a(S_2) - \lambda(S_2))$ is:

$$f_{D_{1,a}, D_{2,a}}(y_1, y_2) = \frac{\Gamma(a)}{a\Gamma(al_1)\Gamma(al_2)\Gamma(a(1 - l_1 - l_2))} \left(\frac{1}{\sqrt{a}}y_1 + l_1 \right)^{al_1-1} \left(\frac{1}{\sqrt{a}}y_2 + l_2 \right)^{al_2-1} \left(1 - \frac{y_1 + y_2}{\sqrt{a}} - l_1 - l_2 \right)^{a(1-l_1-l_2)-1}.$$

By Scheffé's theorem (Billingsely 1999, page 29), it is enough to show that:

$$f_{D_{1,a}, D_{2,a}}(y_1, y_2) \rightarrow f(y_1, y_2) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left\{ -(y_1 \ y_2) \Sigma^{-1} (y_1 \ y_2)^T / 2 \right\}, \quad (3.1)$$

where $\Sigma = \begin{bmatrix} l_1(1 - l_1) & -l_1 l_2 \\ -l_1 l_2 & l_2(1 - l_2) \end{bmatrix}$.

Use Stirling's formula (Wilks 1963, page 177)

$$\Gamma(z) \approx \sqrt{2\pi} z^{z-\frac{1}{2}} e^{-z}, \text{ as } z \rightarrow \infty,$$

where we use the notation $f(z) \approx g(z)$ as $z \rightarrow \infty$ if $\lim_{z \rightarrow \infty} \frac{f(z)}{g(z)} = 1$, to get:

$$\begin{aligned}
 \lim_{a \rightarrow \infty} f_{D_{1,a}, D_{2,a}}(y_1, y_2) &= \frac{1}{2\pi} \lim_{a \rightarrow \infty} \left[\frac{\left(\frac{1}{\sqrt{a}}y_1 + l_1\right)^{al_1-1} \left(\frac{1}{\sqrt{a}}y_2 + l_2\right)^{al_2-1}}{l_1^{al_1-\frac{1}{2}} l_2^{al_2-\frac{1}{2}}} \right. \\
 &\quad \left. \frac{\left(1 - \frac{1}{\sqrt{a}}y_1 - \frac{1}{\sqrt{a}}y_2 - l_1 - l_2\right)^{a(1-l_1-l_2)-1}}{(1-l_1-l_2)^{(1-l_1-l_2)a-\frac{1}{2}}} \right] \\
 &= \frac{1}{2\pi \sqrt{l_1 l_2 (1-l_1-l_2)}} \lim_{a \rightarrow \infty} \left[\frac{\left(\frac{1}{\sqrt{a}}y_1 + l_1\right)^{al_1-1}}{l_1^{al_1-1}} \right. \\
 &\quad \left. \frac{\left(\frac{1}{\sqrt{a}}y_2 + l_2\right)^{al_2-1} \left(1 - \frac{1}{\sqrt{a}}y_1 - \frac{1}{\sqrt{a}}y_2 - l_1 - l_2\right)^{a(1-l_1-l_2)-1}}{l_2^{al_2-1} (1-l_1-l_2)^{a(1-l_1-l_2)-1}} \right] \\
 &= \frac{1}{2\pi \sqrt{l_1 l_2 (1-l_1-l_2)}} \lim_{a \rightarrow \infty} \left[\left(1 + \frac{y_1}{\sqrt{a}l_1}\right)^{al_1} \right. \\
 &\quad \left. \left(1 + \frac{y_2}{\sqrt{a}l_2}\right)^{al_2} \left(1 - \frac{y_1+y_2}{\sqrt{a}(1-l_1-l_2)}\right)^{a(1-l_1-l_2)} \right] \\
 &= \frac{1}{2\pi \sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \exp \left\{ \lim_{a \rightarrow \infty} a \ln v_a \right\}, \tag{3.2}
 \end{aligned}$$

where

$$\sigma_{11} = l_1(1-l_1), \quad \sigma_{22} = l_2(1-l_2), \quad \rho_{12} = -\sqrt{\frac{l_1 l_2}{(1-l_1)(1-l_2)}}, \tag{3.3}$$

and

$$v_a = \left(1 + \frac{y_1}{\sqrt{a}l_1}\right)^{l_1} \left(1 + \frac{y_2}{\sqrt{a}l_2}\right)^{l_2} \left(1 - \frac{y_1+y_2}{\sqrt{a}(1-l_1-l_2)}\right)^{1-l_1-l_2}.$$

Observe that,

$$\begin{aligned}
 \lim_{a \rightarrow \infty} a \ln v_a &= \lim_{a \rightarrow \infty} \frac{1}{1/a} \left[l_1 \ln \left(1 + \frac{y_1}{\sqrt{a}l_1}\right) + l_2 \ln \left(1 + \frac{y_2}{\sqrt{a}l_2}\right) \right. \\
 &\quad \left. + (1-l_1-l_2) \ln \left(1 - \frac{y_1+y_2}{\sqrt{a}(1-l_1-l_2)}\right) \right]
 \end{aligned}$$

Using L'Hospital's rule we obtain $\lim_{a \rightarrow \infty} a \ln v_a$ equals to:

$$\begin{aligned}
 & \lim_{a \rightarrow \infty} \left[\frac{l_1 \frac{-y_1}{2l_1 a^{3/2}}}{\left(1 + \frac{y_1}{l_1 \sqrt{a}}\right)} + \frac{l_2 \frac{-y_2}{2l_2 a^{3/2}}}{\left(1 + \frac{y_2}{l_2 \sqrt{a}}\right)} - \frac{(1-l_1-l_2) \frac{-(y_1+y_2)}{2(1-l_1-l_2)a^{3/2}}}{\left(1 - \frac{y_1+y_2}{(1-l_1-l_2)\sqrt{a}}\right)} \right] (-a^2) \\
 &= \lim_{a \rightarrow \infty} \frac{a}{2} \left[\frac{l_1 y_1}{l_1 \sqrt{a} + y_1} + \frac{l_2 y_2}{l_2 \sqrt{a} + y_2} - \frac{(1-l_1-l_2)(y_1+y_2)}{(1-l_1-l_2)\sqrt{a} - (y_1+y_2)} \right] \\
 &= \lim_{a \rightarrow \infty} \frac{a}{2} \left[\frac{l_1 y_1 y_2 + (1-l_2)y_1^2}{(l_1 \sqrt{a} + y_1)((1-l_1-l_2)\sqrt{a} - (y_1+y_2))} \right. \\
 &\quad \left. + \frac{l_2 y_1 y_2 + (1-l_1)y_2^2}{(l_2 \sqrt{a} + y_2)((1-l_1-l_2)\sqrt{a} - (y_1+y_2))} \right] \\
 &= -\frac{l_2(1-l_2)y_1^2 + 2l_1 l_2 y_1 y_2 + l_1(1-l_1)y_2^2}{2l_1 l_2(1-l_1-l_2)} \\
 &= -\frac{(1-l_1)(1-l_2)}{2(1-l_1-l_2)} \left[\left(\frac{y_1}{\sqrt{l_1(1-l_1)}} \right)^2 + \left(\frac{y_2}{\sqrt{l_2(1-l_2)}} \right)^2 \right. \\
 &\quad \left. + \frac{2y_1 y_2}{(1-l_1)(1-l_2)} \right] \\
 &= -\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{y_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{y_2}{\sqrt{\sigma_{22}}} \right)^2 - 2\rho_{12} \left(\frac{y_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{y_2}{\sqrt{\sigma_{11}}} \right) \right],
 \end{aligned}$$

where σ_{11}, σ_{22} and ρ_{12} are defined in (3.3). The proof follows by using (3.2).